DOCUMENT RESUME

ED 119 727 IR 003 200

AUTHOR	Herman, Elizabeth, Ed.; Byrun, John, Ed.
TITLE	Final Report of the Catalog Code Revision Committee Subcommittee on Rules for Cataloging Machine-Readable Data Files.
INSTITUTION	American Library Association, Chicago, Ill. Resources and Technical Services Div.
PUB DATE	Jan 76
NOTE	58p.; Prepared by the Catalog Code Revision Committee
EDRS PRICE	MF-$0.83 HC-$3.50 Plus Postage
DESCRIPTORS	*Bibliographic Citations; *Cataloging; *Data Bases; Data Processing; Information Storage; Information Systems; Library Automation; Library Standards
IDENTIFIERS	*Anglo American Cataloging Rules; Machine Readable Data Files

ABSTRACT

A committee on rules fur cataloging machine readable data files (MRDF) recommends areas for revision of the Anglo American Cataloging Rules to integrate collections of MRDF into the mainstream of bibliographic control. The 12 working papers which support and detail these recommendations are appended to this report. They include rationales for documentation and control of MRDF and omission of descriptions of physical characteristics from catalog entries. Bibliographic descriptions, medium designation, main and added entries, title considerations, uniform titles, "edition" as an indicator, production and distribution, notes, and summaries are discussed as they pertain to cataloging. The size of file areas in terms of machine readable records and other descriptions is addressed. One report details alternate rules for archival materials, examplars, and other exceptional MRDF. A glossary of terms related to MRDF concludes the report. (CH)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED119727

FINAL REPORT OF THE CATALOG CODE REVISION COMMITTEE SUBCOMMITTEE ON RULES FOR CATALOGING MACHINE-READABLE DATA FILES

JANUARY, 1976

Lawrence Auld
Henriette Avram
Sharon Chapple
Gerry Dobbin
Walter Fraser
Elizabeth Herman (chairperson)
Carolyn Kacena
Judith Rowe
and
John Byrum
(first chairperson)

The body of this report summarizes the recommendations of the Subcommittee. Working papers supporting, detailing and rationalizing these recommendations are supplied as Appendices which are referenced at the point of their first application in the body of the report. Their content is essential to an understanding of the Subcommittee's recommendations and to the preparation of detailed rules. A Glossary has also been prepared which explains the use of various terms related to the field of Machine-Readable Data Files (MRDF), as used in this report. It is recommended by the Subcommittee that these terms be considered for inclusion in the Glossary accompanying AACR-II.

It is the further recommendation of the Subcommittee that several footnotes be provided in that portion of AACR-II covering MRDF (MRDF = singular or plural as context requires)*. One footnote should list frequently-used specific medium terms which are subsumed in the term Machine-Readable Data File; another should list physical characteristics of MRDF which need not be considered for inclusion in catalog records. The Subcommittee feels that such assistance will be necessary for some time to come as the number of catalogers who must deal with MRDF increases as MRDF use becomes more wide-spread.

* (and read as though the words appeared)

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

IR 003 200

FINAL REPORT OF THE CATALOG CODE REVISION COMMITTEE
SUBCOMMITTEE ON RULES FOR CATALOGING MACHINE-READABLE DATA FILES

INTRODUCTION

When the original four members of this Subcommittee met for the first time at the 1971 Midwinter Conference in Los Angeles, they brought very different perspectives to the Subcommittee's charge "to isolate the requisite points of description and recommend methods of description, but not to write rules for cataloging: to differentiate among elements essential to a catalog record and those merely desirable, and to identify those characters subject to easy and frequent change; [and] to look beyond the matter of bibliographic records to explore other areas relating to data files where standardization might develop." Although additional members have subsequently joined the Subcommittee, one of its greatest assets has remained the diversity of backgrounds of its members and of their contacts, which has served to reveal the many facets of machine-readable data files requiring understanding and interpretation before undertaking bibliographic control.

Constituted by direction of the ALA/RTSD Cataloging and Classification Section's Executive Board to its Descriptive Cataloging Committee to form such a subcommittee (subsequently to be transferred to CCRC's purview for the duration of its activities), it was a response to the need increasingly felt in research libraries, to integrate the burgeoning collections of MRDF on their campuses and often in their charge, into the mainstream of bibliographic control. The moment seemed appropriate as the arduous efforts at revising and extending rules for the control of non-book materials (to which MRDF seemed to belong) appeared to be reaching fruition. At a later point, the decision to embark upon AACR-II lent impetus to the effort.

It has been the goal of the Subcommittee throughout its deliberations to deliver final recommendations of greatest utility to libraries and their public by proposing rules as nearly as possible parallel to and certainly compatible with the body of rules presently operative or intended to be operative in AACR-II. This has resulted in decisions which seem at first to go against some popular conceptions of the needs of potential users. For this reason, the Subcommittee wishes to emphasize that these proposals seek to achieve the purpose the catalog

traditionally serves: that of informing the user that a work by a given individual or group of individuals, with a certain title, possibly related to other works, titles or editions, and upon a particular subject is available for use. The Subcommittee envisions users of two general types, those who have some sophistication in the area of MRDF and those who do not. In either case, the requirements of the medium are such that a completely detailed description of a MRDF would entail such extensive catalog records as to confuse rather than assist. The Subcommittee has no doubt that whatever catalog record is created, backup details and user assistance must be available before a MRDF can be used. It is the Subcommittee's recommendation that such additional information needs be recognized in the introductory remarks preceding the special rules. The envisioned information points will provide the additional documentation that a knowledgeable user can examine and will provide the guidance a less-knowledgeable user requires. In either case, the catalog record would have served its purpose in identifying a potentially useful work. (Appendix A: Documentation and control of MRDF).

A further divergence from other proposals for description of machine-readable data files is the Subcommittee's decision to avoid detailing physical characteristics of MRDF, because of the ready convertibility of the data file's physical form or format. For example a MRDF can be successively stored on cards or tape or a disk, or its storage density within the same medium may be changed at the convenience of the processing center. This would render the catalog records containing such information inaccurate or extremely costly to maintain. (Appendix B: Omission of descriptions of physical characteristics from catalog entries...)

Another distinction between machine-readable data files and most other media regularly dealt with in AACR is found in the sources of information from which a description must be derived. While the item in hand is usually the primary source of information in cataloging, a machine-readable data file often affords little that provides useful bibliographic information. In contrast to that of some other non-book materials, the MRDF container is frequently not labeled at all or it has a casual and unreliable label. For these reasons, the most useful and reliable sources of information are usually found in documentation external to the MRDF and its container. Such sources vary from announcements of research being conducted utilizing machine technology, distributors' announcements of availability, codebooks accompanying the file, even to verbal transmission from the originator or depositor of the file. While the primary source must remain the item itself, the Subcommittee in recognition of reality, recommends that rules provide for derivation of bibliographic descriptions from secondary sources, until such time as the incorporation of such information within a MRDF becomes regularized. An enumeration of possible sources of information available and some guidance to catalogers in utilizing these sources (but not by a priority list) should be provided in an introductory passage. (Appendix BB: Sources of bibliographic descriptions)

The medium designator Machine-readable data file, is proposed by the Subcommittee as specific enough to distinguish the medium, broad enough to include the variations within the medium, and of sufficient currency and clarity to be easily recognized by catalog users. (Appendix C: Medium designation)

An ISBD(NBM) to include MRDF is presently under study by another group, and will presumably be compatible with existing ISBD concepts. On the basis of the ISBD(M) model, the Subcommittee does not foresee any insurmountable problems in applying an ISBD format to MRDF if the special terminology and characteristics are accommodated. These special features will be developed in the following paragraphs, but the overall description of MRDF appears to be reasonably accommodated by the basic precepts of ISBD.

RULE RECOMMENDATIONS

Entry: As the creation of the content of machine-readable data files differs very little from that of printed materials, the Subcommittee recommends applying the rules of entry for monographs (see note on Serials, below) as outlined in AACR, chapter one, rather than any special non-book rule. Added entries should be made in accordance with a liberal interpretation of AACR 33. (Appendix D: Main and added entries).

Transcription of the title should follow the present course of ISBD: Title [medium designator] = parallel title : subtitle : other title information. As most MRDF titles will be taken from secondary (but acceptable) sources external to the file, the Subcommittee recommends that these be transcribed without brackets. The source of the title should be specified in a note. Brackets should enclose titles created by the cataloger. (Appendix E: Title considerations). Statements of responsibility found within the file or in its appropriate documentation should follow.

Application of uniform titles: The Subcommittee recommends that uniform titles be applied to MRDF when they are works which would receive uniform titles in other media. Application will vary in libraries according to their size and scope, just as it does for other media today. In cataloging MRDF which are the same file in terms of content and which have been given no title but are known by a variety of local titles which vary from location to location, user to user, etc., the Subcommittee recommends caution in choosing a uniform title from amongst them, and urges rather that no uniform title be selected until such a file is cited consistently in literature, or the author has designated a title later, or formal distribution has resulted in a stable title citation. Appropriate notes can relate one title to another if necessary. (Appendix F: Uniform titles)

An Edition statement should be based on the presence of an edition statement in the MRDF or its documentation, a difference in the content of the file, a difference in the count or size of the logical records, or the naming of a different editor or compiler. Differences in the program language of a machine-readable data file or, in text files, a difference in the text input, also indicate another edition. Physical differences, copying or distribution date variations do not in themselves indicate another edition. (Appendix G: Edition). In keeping with ISBD practice, a statement reflecting responsibility for an edition should follow if such information is found in documentation being used as the basis of the catalog record.

A Production and distribution area should consist of a place of production, the name of the producer, the date of production, the distributor and the distributor's location. The usual provision that information pertaining to the cataloging agency's country be recorded in addition to that of a first listing of that in another country, would hold for MRDF. Care must be taken to avoid confusing the act of copying a machine-readable data file with its production or distribution. (Appendix H: Production and distribution)

Size of file area: For reasons indicated in the introduction of this report and elaborated in Appendix I: (Size of file area), the Subcommittee has settled upon the number of logical records in a MRDF as a serviceable description of the size of file. For program files, the number of statements coupled with the programming language is proposed [300 COBOL statements]. Size of file should be omitted for Object Programs and a note supplied in the note area of the nature of the program and the machine on which it runs [Object program (IBM 360/40)]. In cases of multi-file works, comparable to book sets for example, the size of file area would state the number of files and the number of logical records in each [3 files (1900, 456, 640 logical records)]. If the catalog record of such a multi-file work includes a contents note, the logical record count of each constituent file may accompany its title. The size of file area would then state only the number of files.

The presence of accompanying material should be introduced by a / and contain a descriptive term and a numerical description if appropriate [/ codebook (320 logical records) or, / accompanying documentation]. More elaborate description when necessary may be found in the note area.

An alternative rule for archival material, exemplars and other exceptional machine-readable material is provided. (Appendix J).

Series statements for MRDF can reasonably follow those for other media when such are determined for AACR-II. (See note on Serials below).

Notes Area: Notes comparable to those used in other media to elaborate some element in the preceding descriptive areas are similarly warranted in the case of machine-readable data files. Additionally, certain notes relating to MRDF particularly may be made: source of title found in documentation; content changes resulting in a different printout; the source of a copied MRDF if provenance is significant; extensions of size of file or details of accompanying material; object program note. A contents note for multi-file works, particularly those whose constituent files have descriptive titles, may substitute for a Summary. (Appendix K: Notes).

A Summary is to be preferred to a contents note except as observed above and is recommended for machine-readable data files, a medium which precludes examination of the material itself. (Appendix L: Summary).

ISNs when assigned should be noted. The ISBN which may be assigned a codebook or some item in accompanying materials should be recorded in this area with an indication of the item to which it has been assigned.

SERIAL MACHINE-READABLE DATA FILES

The Subcommittee has withheld extended consideration of serial MRDF because of the on-going general discussion and the indecision of the library community on rules of entry and description, discussions not concluded at the time these recommendations were compiled. However, serial MRDF have sufficient in common with serials in other media - in addition to the basic requirement of seriality - to render them amenable to a general rule. The Subcommittee does caution that the work being cataloged should be that work as issued. The flexibility of machine-readable data files permits a great many optional treatments of the individual issues which may result in completely different cumulations and holdings in any given location. Description of such resultant files should be in addition to the basic catalog record or be a part of the additional information available at a secondary information point.

Proposed footnotes in AACR:

Storage media subsumed in the designation machine-readable data file.

It is recommended by the Subcommittee that these terms be listed in a footnote or introductory note in the text of AACR with the caveat that such storage media so subject to change in the data processing milieu need not be noted in catalog records. Otherwise, catalogers seeing such terms so frequently will feel that they must record them.

Aperture card, cartridge, cassette, computer tape, data cell, direct access storage device, disk, disk pack, drum, edge-notched card, Hollerith card, machine-readable identification card, magnetic disc, magnetic tape, punched card, punched paper tape, tape, tape reel.

Commonly listed terms found in various contexts and applying to physical characteristics to be excluded from catalog records (but not necessarily from other levels of records).

Again the Subcommittee recommends this information be the subject of a footnote to put a cataloger's mind at rest when faced with the decision to include in or exclude from the catalog record such information.

ASCII, bar code, binary-coded decimal, bit, bits per inch (bpi), block, byte, channel, character, characters per inch (cpi), column binary, core, cylinder, EBCDIC hexadecimal, octal, optical bar code, parity, physical record, track, volume.

DOCUMENTATION AND CONTROL OF MACHINE-READABLE DATA FILES (MRDF)

The following remarks are offered to clarify the assumptions underlying the recommendations of the Subcommittee on Rules for Cataloging Machine-Readable Data Files.

The Subcommittee agreed very early in its discussions that in order to best satisfy the needs of users, simplify the job of the librarian and provide a sufficiently flexible system for the many ways in which data libraries are organized and administered, it would be necessary to think in terms of a multi-level system of documentation and control.

The discussion which follows provides an example of how such a system might be implemented. Since it is only for the user who wishes to actually access the MRDF that we would have to provide full documentation and a physical description of the medium on which the data are stored, the following approach to bibliographic control, involving four distinct types of records, seemed a practical solution: standard catalog entries; data abstracts (or data description forms); content documentation or codebook; and a record of the physical and logical characteristics of the data file. These records or any variants of them would be interrelated just as are the complex records necessary for utilization of other types of library materials such as incunabula, manuscript collections and serial publications. However, it is not the intention of the Subcommittee to concern itself with standards for anything but the catalog record, since only the catalog record will necessarily be maintained by the library.

Standard Catalog Entry

It is appropriate that academic and research institutions would want to record and provide access to files of data in machine-readable form in the public catalogs of their libraries where entries already appear for other forms, such as books, pamphlets, serials, and also often for microforms, recordings, and motion pictures.

The insertion of standard bibliographic records -- perhaps someday generated by a central agency such as the Library of Congress -- into the library(s general catalogue is a primary method of accessing and controling an institution's machine-readable data holdings. As such, it also offers a handy means by which to advertise acquisitions of this type of material. Perhaps the public catalog will produce the side benefit of easing somewhat that circular problem confronting all data librarians (and to a greater or lesser extent all archivists and reference librarians) that of the underutilization of existing informational resources.

More importantly, the development of standards for cataloging MRDF also paves the way for the emergence of a national union catalog for machine-readable data files. Already well-developed for book and booklike materials, including microform masters and manuscripts, such union lists are based on reports from a large number of participating institutions which describe holdings according to generally accepted cataloging rules. A union list of machine-readable data files would enable an institution

to alert a far greater constituency than that attached to it to the availability of specific files and would enable the individual researcher to locate easily relevant files beyond those held by the institutions within his immediate reach.

Standard catalog entries for machine-readable data files would of course be compatible with those for other library materials.

The Data Abstract or Data Description Form

A data abstract should include all of the elements necessary for transcribing the imprint equivalent onto the catalog record, (author, title, place and name of producer, place and name of distributor, date of distribution, etc.) as well as information appropriate for the note area, and information useful in determining subject classification. Such information would include where relevant: dates of study or data collection; date of file production; the universe to which the data pertain; method of data collection; number and definition of logical records; number and type of fields or variables contained in records; structure of file; condition of the data; reference materials, related publications and sources of further documentation; relevant programs; and restrictions, if any, on the use of the file.

In addition there would be a summary which would indicate briefly and generally the content of the file including major subject areas covered; any special features; clarification of title or dates, if necessary; and, when applicable, a statement as to how often the file is issued or updated.

Other items of information might also be included if they contributed to the potential user additional information for making an intelligent decision about the relevance of the data file for his research or information needs.

This record would normally be a one-page summary which could be entered into a loose-leaf binder or it might exist in machine readable form and be accessed from a time-sharing terminal. These records should be produced by the individual or archive with primary responsibility for making the data available.

The National Bureau of Standards has prepared a form, "Software summary for describing computer programs and automated date systems," as a Federal Information Processing Stand (FIPS Pub 30, 1974) which serves as one model.

Content Documentation or Codebook

This would contain detailed descriptions of each individual field, variable or date element contained in a logical record which make up the file including its content, location and size. Information about missing data, special flags to indicate parts of speech, special codes used for non-standard characters, and the like, would be mentioned. In the case of survey data, for example, a codebook should contain the location and width (deck and columns) for every variable or question and the code number and description for every response. Attached to this codebook

might also be the original questionnaire, instructions to the interviewers and/or coders, and any other material which might be of use to the potential user. The quantity of documentation for a given data file will depend on the nature of the data, the variety and number of variables, and the presence of related published materials. Statistical or coded data would normally require more documentation than text.

Such documentation would hopefully be produced and distributed with the data file itself and could be either in hard cover or in machine readable form.

Record of Physical and Logical Characteristics of the MRDF

This record should specify volume number(s) (e.g., tape, disk or drum); file number(s); character density (e.g., 800 characters per inch); track (e.g., 7 or 9); parity (odd or even), if relevant; recording mode (e.g., EBCDIC, BCD or binary); record format (e.g., fixed block or variable block), logical record length; block size; number of physical and/or logical records; data set name; tape label; header label; and volume identifier, etc. There should be a separate record for each copy of the file and for each volume if the data set comprises more than one volume.

The record of physical and logical characteristics would always be produced locally, preferably by the data library or computer facility at which the data file is to be physically stored.

OMISSION OF DESCRIPTIONS OF PHYSICAL CHARACTERISTICS FROM CATALOG ENTRIES FOR MACHINE-READABLE DATA FILES

The Subcommittee on Rules for Cataloging Machine-Readable Data Files recommends that descriptions of physical characteristics be omitted from the catalog entries for data files. Such information is not necessary for bibliographical identification of particular files. Moreover, the ease with which these properties are changed and the specialization required to describe them make it also impractical and undesirable to incorporate details of physical characteristics into catalog records for machine-readable data files.

The degree of specialized knowledge required to include a complete physical description in the catalog record for a data file represents a practical barrier to such an undertaking. In order to provide a physical description which would be complete enough to be helpful to the catalog user, a cataloger would need to have a thorough grasp of the various storage media, the total extent of encoding possibilities, and the multitude of types of equipment in use in all parts of the world at all times. Programmers, systems analysts, and hardware experts do not attempt this kind of specialization; they depend on a combination of experience, consultation, and trial and error. To supply a creditable physical description, then, a cataloger would have to become an outstanding specialist in a non-cataloging field.

More importantly, the frequency and casualness with which the physical characteristics of a data file may be -- and very frequently are -- altered without any consequent change in the informational content greatly diminishes the likelihood of an accurate physical description in the catalog even if one were attempted. The physical forms in which machine-readable data files are housed are subject to a remarkable amount of change even though the content may remain static. As a data file is used by one researcher or another and in one facility or another, it may be convenient or even necessary to alter the physical form in order to use the file with the available equipment. Punched cards, a bulky but otherwise relatively low cost storage medium, are likely to give way to magnetic tape as closet space is exhausted or when a file is transported to another location. During processing it is often preferable, sometimes mandatory, that the file be transferred to disk or data cell. Or Storage density within the same medium may be changed. Or the coding system may be changed. In some facilities, the physical format is dictated by policy and may be altered without the user being either consulted or notified. Keeping track of such volatile details as a part of the catalog record is an impractical enterprise; it places a great workload on cataloging staff, and rarely would catalog records be up-to-date and therefore accurate in their physical descriptions. Moreover, the actual data content in the file need not change as the file is reformatted and transformed for reasons of local convenience.

Nor is it essential to the purposes of descriptive cataloging that a statement of physical properties appear on catalog entries for machine-readable data files. Unlike books, maps, motion pictures, etc., machine-readable data files exist independently of the particular materials on which they are stored. Therefore, their physical properties, being variables, do not assist in distinguishing data files as bibliographic entities and would not necessarily facilitate identification of particular files if included on catalog records.

It is true that actual manipulation of a specific data file, once a decision to consult it has been made, will require familiarity with the physical properties of the file as it is currently stored as well as with the properties of the machinery and programs available. For this reason, data abstracts, codebooks, or data sheets which are sometimes prepared for distribution with a file may often include a physical description. However, the file does not necessarily remain in this form, and, within the local computer facility, it is customary to keep track of the current physical status of files, even if only informally. In addition, most users of data files are not themselves concerned with the physical characteristics of a file; such matters are usually left to computer center personnel so long as the information is available through this source. It is therefore appropriate that the details of physical characteristics be available on records maintained by the agency responsible for actual utilization of these files, not on the catalog entry.

As a matter of record and information, a comprehensive list is attached detailing those physical characteristics which would be omitted as a result of this recommendation. Undoubtedly, there are obscure and new characteristics which have been overlooked, but this list is indicative of the Subcommittee's intent. At a later date the Subcommittee will prepare a recommendation for an alternative to a physical collation for machine-readable data files which will indicate to catalog users the relative file size (i.e., quantity of data). Consideration is also being given to the type of arrangement (or file structure) employed in the file.

List of Variable Physical Characteristics of Machine-Readable Data Files*

1. Medium

   a. Magnetic tape (plastic or metallic)
   b. Punched card (21, 45, 80 or 160 column)
   c. Punched card with magnetic tape strip
   d. Ledger sheet with magnetic tape strip
   e. Aperture card
   f. McBee cards
   g. Punched paper tape
   h. Disk pack or fixed disk
   i. Bar-coded document
   j. Mark-sensed score sheet
   k. Magnetic Ink Character Recognition document
   l. Optical Character Recognition font document
   m. Machine-readable identification card
   n. Microfile (e.g., MIRACODE)
   o. Data cell

2. Type

   a. Nine channel (tape)
   b. Seven channel (tape)
   c. Cassette (single reel); cartridge
   d. Cassette (double reel)
   e. Cassette (closed loop)
   f. MTST (tape)
   g. 80 column (card)
   h. 90 column (card)
   i. Stub (card)
   j. IBM System 3 (card)

3. Size or quantity

   a. Reels - number and diameter (tape)
   b. Size (cassette)
   c. Length (tape)
   d. Width (tape)
   e. Number of columns (cards)

4. Code

   a. EBCDIC
   b. Hollerith
   c. Binary Coded Decimal
   d. Octal
   e. Hexadecimal

---

*The attempt was made to make this list as inclusive as possible covering both common and little known characteristics. However, the list remains only illustrative and it is not to be assumed that the omission of a physical characteristic from this list implies approval of its inclusion on the catalog record. All of these items where applicable belong on a Record of Physical Characteristics.

4. Code (continued)

   f. ASCII (64 character)
   g. ASCII (128 character)
   h. Binary
   i. Column Binary
   j. Biquinary
   k. XS3

5. Recording characteristics

   a. Density (bits per inch)
   b. Density (characters per inch)
   c. Parity (odd or even)

6. Volume and file identification

   a. Volume identification
   b. Labeling information
   c. File number or track locations
   d. DSNAME (data set name)

SOURCES OF BIBLIOGRAPHIC DESCRIPTIONS

In general, the bibliographic characteristics of machine-readable data files cannot be determined from the materials themselves. At this stage of their development, the various species of machine-readable files--survey, program, historical, textual, archival--share a common lack of internal indicia sufficient to identify the distinguishing features of specific editions of specific titles. (Nor do current ANSI standards for labelling always provide for incorporation of a file's full title, let alone an interior record of its other bibliographic details.) In fact, like many incunabula, MRDF are often issued without any definite title assigned to them.

Most of the information normally presented in the body of the entry of catalog records--details concerning title, authorship, edition, publication--as well as information more suited to note position, then, will necessarily be derived from sources external to the MRDF itself, whether in its machine-readable format or in hard copy as printout. Descriptive cataloging elements, however, usually may be formulated on the basis of secondary sources. These include: (1) descriptions provided by the collector of the data, if the file is survey or historical in nature, or by the editor, if the file consists of textual matter; (2) descriptions provided by the agency (archive) which produces and distributes the machine-readable file; (3) other published descriptions of the file, whether those produced by an abstracting service, or those contained in a scholarly publication based on or related to the file; (4) descriptions included in such MRDF documentation as codebooks; and, (5) descriptions which the cataloger himself may undertake to supply. It seems to be true that, in characterizing any given MRDF, secondary sources may sometimes differ in terminology by which they describe the file and in the accuracy of the information they supply.

The fact that MRDF are issued without any title-page equivalent and may be described variously by the secondary sources from which the catalog record will be constructed raises at least two questions. First, should the cataloger be required to identify somewhere in the entry the sources from which his description resulted? As a general guide, the Anglo-American Cataloging Rules states: "The source of information contained in an entry need be specified only when the information is questionable or the source is unusual"* (p. 190). No single source has yet emerged

---

*When dealing with descriptions of book and book-like materials, the rules favor identification of sources exterior to the item cataloged. With the other media, the rules vary on this matter. Thus, when supplying the title of a phonorecord, the cataloger is instructed: "If the title is found on the phonorecord being cataloged, no note is needed; if, on the other hand, the title is transcribed from an album cover, container, etc., a note giving the source of title is made" (p. 324). Chapter 12 rev. prescribes a note of source of title if the source is not the work itself, material accompanying it or the integral or unifying container. (229.2B p.25)

which might be recommended as providing complete and accurate bibliographic information about data files at this stage of their development. Descriptions provided by the creators of files (and sometimes also those produced by processing archives) are frequently -- but certainly not always -- reliable sources of cataloging data. Unfortunately, these descriptions do not necessarily contain all the information which the cataloger may be required to supply; nor may he reasonably expect such descriptions to routinely accompany files which he must catalog. Until labelling standards are revised to encourage internal identification, and files actually begin to incorporate adequate labels, this situation is not likely to change. It is, therefore, suggested that external sources from which cataloging data for machine-readable files is derived be revealed to some extent in a note. A full enumeration of all sources utilized to produce entries for data files would often yield records which contained documentation that exceeded in size the remainder of the description. To keep entries relatively brief and simple while revealing that the information contained in them came from external sources, the cataloger should normally note only the primary source of his description. Usually the source cited should be that from which the cataloger takes the title of the file.

Secondly, should the rules for data files specify an order of preference to guide the cataloger in resorting to sources from which he will prepare entries? The Anglo-American Cataloging Rules may establish for the cataloger a fairly firm order when dealing with the variety of possible sources of bibliographic information for more traditional library materials like serials and music.

As for sources of bibliographic characteristics to be recorded in the body of the entry, the cataloger could depend on the machine-readable data file itself only when it includes a title-page equivalent in the form of an adequate interior label. If this condition is satisfied, the data file will constitute the cataloger's primary input of bibliographic information. In the absence of such a label, the cataloger will determine the file's bibliographic characteristics from whatever descriptions that the body from which the library acquires a data file may have provided. This source may be the creator of the file, the editor of a pre-existing file, the distributor of a file, or a donor; and, as indicated above, the adequacy of such descriptions should not be taken for granted. In lieu of materials prepared to accompany a machine-readable data file -- or in addition to them -- the cataloger may find it necessary to seek out other published descriptions of the file, such as those found in certain reference works or in articles and monographs based upon the file. Because a file's container is never an integral part of the data stored within it, and because such containers rarely provide what would be considered bibliographic information, the container is generally to be disregarded as a source for the catalog record.

With regard to obtaining information concerning the size (and possibly also the condition) of the file, the machine readable data file itself becomes the preferred source. It is often true of MRDF that they are issued in an imperfect and/or incomplete state, that lacunae are frequently supplied

as new and more data becomes available, that material is revised as defects are noticed, and that the file is reorganized to suit some special purpose. Therefore, any generalized descriptions pertaining to the content of a given file -- especially to the amount of data contained within it -- may differ from the actual condition of the copy being cataloged, whether the version described in secondary sources is considered perfect or not.

Thus, it is advisable to verify the size of file through actual counts of data, if possible. Nevertheless, those actually responsible for cataloging MRDF may be physically removed from the computer facility which houses them. Or, there may be no provision of computer time to the cataloger for verifying data files. Consequently, the rules for describing them, although they may express a preference for the MRDF itself as a basis of establishing the size and condition of the file, should allow as an alternative source for this field the description(s) from which the cataloger is drafting the body of the entry. Whenever secondary sources alone supply the size of the file, the cataloger could note that the size of file had not been verified.

Generally, then, the following -- if not construed to represent a fixed order of preference -- constitutes a recommended sequence of sources to consult in deriving the bibliographic characteristics of particular machine-readable data files: (1) the data files themselves, their containers generally to be disregarded; (2) documentation prepared to accompany the files; and (3) other published descriptions. In an introductory section, the rules for cataloging machine-readable files should enumerate the possible sources of description and provide guidance in their use. (For a table which suggests the relevance of such sources to specific bibliographic fields, see page following.)

In summary, it is recommended that the rules for cataloging machine readable data files:

(1) explicitly recognize that bibliographic descriptions for this media will largely be derived from secondary sources until such time as labelling standards are revised to require the incorporation of a title-page equivalent;

(2) require that the external source from which title is taken be specified in a note; and

(3) include in an introductory section an enumeration of the usual sources of information for bibliographic descriptions of data files and provide guidance in using them.

RELEVANCE OF SPECIFIC SOURCES OF BIBLIOGRAPHIC INFORMATION FOR SELECTED CHARACTERISTICS OF MACHINE-READABLE DATA FILES

SOURCES:	Creator of file	Producer or Distributor	Other documentation	Data in machine-readable form	Data in hard copy	Other published descriptions
Bibliographic characteristic						
Title(s)	X*	X*	X	0	0	X
Production	X*	X*	X	0	0	X
Release/ Distribution	X	X*	X	0	0	X
Dates data collected/ period covered by file content	X*	X	X	0	0	X
Data file produced/ released	X*	X*	X*	0	0	X*
Size of file	X	X	X	X*	X	X
Series	X*	X*	X	0	0	X
Restrictions on use	X*	X*	X*	0	0	X*
Condition of file	X	X	X	X*	X	X
Summary of contents	X*	X*	X	0	0	X*

Key: X* - usually a primary source
X - other likely source of information
0 - not generally applicable

MEDIUM DESIGNATION

In its consideration of what has come to be referred to by the Subcommittee as "machine-readable data files," it has been implicit that a medium designation should be applied to the titles of such cataloged materials--primarily in order to distinguish them from the same works in other media, particularly the printed.

The point has already been developed (cf. DCC/MRDF-2) and should be reiterated here that this medium entails a characteristic which differentiates it from other media generally cataloged in that its physical format is readily changeable in the course of utilization. This consideration should be recognized in the choice of a designation for the medium as a whole: The terminology used for the designation should be sufficiently general to accommodate this characteristic but sufficiently delimited to exclude vagueness or confusion with other media. Further, the designation should be compatible with natural language rather than so contrived as to seem artificial to users.

Suggestions have included descriptors introduced by the word "computer" and incorporating the name of a particular storage device such as "computer tape," "computer disk," etc., which fall into the category to be avoided because of the unpredictable convertibility during utilization of the material as suits processing convenience. A file which resides now on tape may later be transferred to disk. Thus, use of such designations would result in the need for uneconomical recataloging or, especially in printed union catalogs, in inaccurate records. Further, the word "computer" is too limiting, as machines other than computers may be used in manipulating the material.

Frequently-heard designations are those introduced by the word "data;"* "data record," "data set," "data file," "data base," "data bank," etc. To many these terms convey a sense of size, a "data item" being the smallest unit, and "data base" or "bank" implying the largest accumulations. Between these extremes, "data set" and "data file" are sometimes used interchangeably, but "data file" is more unambiguously defined as a collection of related records to be treated as a unit, while definitions for "data set" vary according to computer languages, glossaries, and individual usage. However, any designators which do not take into account the means of access to the information do a disservice to the catalog user, as any of the terms introduced by "data" could conceivably apply to information in another medium.

*In this context, "data" may be defined as "the quantities, characters, or symbols on which operations are performed by computers and other automatic equipment, and which may be stored or transmitted in the form of electrical signals, records on magnetic tape or punched cards, etc." It is sometimes felt that "data" is not an appropriate word by which to refer to textual and bibliographic files; such a view takes a narrower interpretation of the term than is warranted by pertinent dictionaries and glossaries. In general, these equate data with "all the facts, numbers, letters, and symbols that refer to or describe an object, idea, condition, situation...."

This brings us to designations introduced by the words "machine-readable," a term which indicates the nature of the processing device and the sensing process. Possibilities include "machine-readable records," "machine-readable media," and "machine-readable data files," etc. While "computer" is too limiting a term to designate the devices through which these materials can be utilized, "machine" is more acceptable in terms of inclusiveness. The term "read(able)" in data processing connotes that data is in a form which can be sensed, retrieved and/or moved by machine. Of the examples above, "machine-readable data file" is to be recommended as indicating that the material cannot be used without certain equipment as well as characterizing its form. "Data file" encompasses the "item" - "record" sequence while stopping short of the collection of units which form a "data base" or "data bank."

The Subcommittee recognizes that the term "machine-readable data file" is longer and more cumbersome than other medium designators. But, after considering at length all alternatives, it reached the conclusion that no other descriptor offers equal precision and clarity. To eliminate "machine-readable" is to subtract the unique character of the medium. To substitute entirely another term for "data file" is to reject the most familiar medium-oriented generic term. In particular, such alternatives as "record(s)" (e.g., "machine-readable records") and "medium" (e.g., "machine-readable medium") are particular inappropriate because of special connotations of these words to users of the materials; as noted above, "record" refers to an element within a "data file," while "medium" or "media" characterizes the material used to carry the file rather than the organizational unit which contains it. The Subcommittee also gave consideration to reducing the length of the designator by using "data" (e.g., "machine readable data") or "file" (e.g., "machine readable file"), but decided that only both components together--with the implication that data reflects the content, file the organization of the content--prove to be fully descriptive. Moreover, the possible redundancy of using "data" and "file" rather than "data" or "file" is counterbalanced by general usage and understanding of the full term.

The Subcommittee, therefore, recommends the term "machine-readable data file" as one which seems to satisfy the needs of the medium for a designator--neither too general nor too specific, and reasonably self-explanatory to a cross section of catalog users.

References:

American National Standards Institute. American national standard vocabulary for information processing, 1970.
Jordain, P. B. Condensed computer encyclopedia, 1969.
Rodgers, H. A. Funk and Wagnalls dictionary of data processing terms, 1970.
Sippl, C. J. Computer dictionary and handbook, 1966.
Weik, M. H. Standard dictionary of computers and information processing, 1969.
Meek, C. L. Glossary of computing terminology, 1972.
A Supplement to the Oxford English dictionary, 1972.

MAIN AND ADDED ENTRIES

We have seen in other discussions of Machine Readable Data Files emanating from this committee that the physical form which at the outset seems to distinguish MRDF so sharply from printed publications, actually has less significance in cataloging considerations than the intellectual content of the files, and further, that this content frequently bears a close resemblance to the content of a printed work. In fact, the content of MRDF may be the identical intellectual content in the other medium. It therefore seems entirely logical to recommend that the main entry of MRDF be determined on the basis of the same criteria which are applied to printed materials. Further, responsibility for the intellectual content may be the area in which the facts are most readily obtainable in this field of as yet sketchy bibliographic control.

In many instances it is the investigator who has gathered the information for processing and analysis, and who goes on to the publication of conclusions based on the data in the file, who may be deemed the agent responsible for the intellectual content of that file and consequently the candidate for main entry. In the instances of construction of concordances or of definitive editions of an author's works as a result of machine comparisons, the author whose work is under purview will be the entry for the file itself (consisting largely of the author's text), although the compiler(s) of the concordance might be the entry for the published work (AACR 19). Often a corporate entry will be found for MRDF: an agency of government, a committee or research group with its own name. The evidence in the file and in its documentation must be assessed as is the evidence surrounding a printed item, and the basic rule for entry applied.

However, this concept of MRDF entry as consonant with that of printed materials is not everywhere accepted amongst those agencies and individuals who for one reason or another have sought to deal with some sort of cataloging of MRDF.

In Non-book Materials Cataloguing Rules, prepared by the Library Association Media Rules Committee (of Great Britain) a basis of adherence to AACR rules of entry is presumed; descriptive rules are provided for those media considered as "non-book" by the committee. The rules were brought forward without addressing Machine-Readable Data Files in order to allow more time for consideration of proposals by Ray Wall who "constituted a one-man working party," together with those which this subcommittee was just then formulating. However, neither the preliminary report nor the draft rules prepared by Dr. Wall directly deal with the concept of main entry. His outline seems to suggest Title as entry, although in a paragraph headed "Credits" he proposes that the name of the organization holding and responsible for processing the file be considered the "author." He considered for cataloging only the "main or master files in respect to a set of data," in part due to the inconstancy of physical format of the content. This committee, however, has attempted to provide guidance for the preparation of catalog entries for any MRDF in a manner useful and usable in multimedia catalogs in a range of libraries, rather than a description of an archetypical file.

In Non-book Materials; the organization of integrated collections (Weihs, Lewis, Macdonald; Canadian Library Association) the rules for MRDF have been based on the early reports of this committee, and do not pursue the matter further than those considerations had at the time of its publication. However, the general remarks on entry refer to AACR.

The Association for Educational Communication and Technology's AECT Standards for Cataloging Nonprint Materials, 4th edition, will substantially alter the oversimplified approach of earlier editions in addressing MRDF. The blanket instruction to enter under title has been abandoned in favor of the basic rules of entry found in AACR, in recognition of the parallels between print materials and MRDF.

Articles in current journals which have occasionally dealt with some aspect of bibliographic control of this medium have usually displayed a pragmatic approach, frankly selecting an element of organization useful to the particular public the catalog serves. Not infrequently this provides a primary arrangement by subject under which a title, actual or descriptive, heads the further description of the file. These arrangements are usually found in catalogs or lists of exclusively this medium and have not had to grapple with the impositions of a multi-media catalog constructed on the basis of the AACR and previous codes.

It may be noted that the latest revision of AACR chapter 12 has also accepted the concept of entry to be determined in accordance with the basic rules set forth in Chapter 1 of AACR, so that this subcommittee substantially concurs with decisions arrived at independently by those formulating rules for other media. Thus, it seems a logical and useful proposal to advocate applying the basic principles of AACR to MRDF to the extent that they are pertinent.

Certain added entries are prescribed as corollary to the rules governing choice of entry (AACR, Chap. 1): a joint author or editor, the editor of a work which receives a title entry, sponsors whose responsibility is more than financial, and in decisions based on rule 17 (works of corporate authorship) for the personal author or corporate body, whichever comes in second in the contest for main entry. Others are discussed in AACR 33 which provides for added entries for corporate bodies openly named as sharing responsibility for the work if such association extends beyond that of an issuing body or financial underwriter (33G). However, 33H goes further to support added entries for persons or bodies that have a relationship to the work, considered important for retrieval purposes. 33K considers for added entry works closely related to the work in hand, which would seem to cover related published texts, summaries, abstracts, etc., citation of which might, if needed, aid identity. 33L appears applicable to archives and repositories which are often the sources of duplicated files and which might seem a very important indicator to a catalog searcher. Added entries for Titles, Series as necessary, and analytical entries if warranted, are also provided for in 33.

The AACR provisions for persons (in 33) are equally hospitable, providing possible added entries for agents of responsibility not chosen for main entry, collaborators, writers, project directors, editors, compilers and translators. The pervasive principle to be applied to any added entry is the utility of that entry to user retrieval in the catalog. In determining this, attention must be paid to the documentation surrounding the file, the importance of persons or groups of persons (corporate or otherwise) represented in this documentation as participating in the process of creation and production of the file; to the provenance of the file; and to the related materials in relation to which the file may best be known. Optional added entries (those not prescribed in the rules for main entry in chapter 1, AACR) should be made with discernment in view of the multiplicity of names and relationships which can be associated with a given MRDF.

Appendix F

TITLE CONSIDERATIONS

The title page of a Machine-Readable Data File may be defined as an internal user label or its substitute preceding or within the file itself. Further, a print-out or CRT image of the same may be considered the equivalent for cataloging purposes. At the present time such labels containing bibliographic information suitable for the derivation of cataloging copy are infrequently found, but publication of standards and growing recognition of the utility of formal identification of MRDF should result in increasing use of the label for regularized identification. Therefore, just as for monographic publications the title page or its substitute within the volume is the primary source of title for cataloging purposes, so for MRDF the internal user label or its substitute within the file should constitute the primary source of title information.

In Appendix BB entitled Sources of Bibliographic Descriptions, further sources external to the file being cataloged which were recommended as fertile ground for discovery of bibliographic details were listed as "documentation prepared to accompany the files" and "other published descriptions". Serious reservation regarding the reliability of the container as a source of information was indicated. The source of the title, it was advised, should be specified in a note. The intent of these recommendations is to encourage utilization of reasonably reliable documentation, that which gives some assurance of forethought, of intent to document. These then are proposed as the source of the title proper: the title page equivalent of the file itself when satisfactorily utilized, or the file's most reliable documentation, this source to be recorded in a note for recognition and comparison with catalog records of machine-readable data files of similar description.

Transcribing the MRDF title

Once the basis of description is established, the title, parallel title(s) and other title(s) ascertained should be transcribed as nearly as possible in accordance with standard cataloging procedures in order to facilitate the interfiling of these records into a multi-media catalog. Despite the difference in physical format, the similarities in terms of content between MRDF and printed materials are greater than their dissimilarities, and the approach to cataloging can reflect these similarities. The rules covering the transcription of titles of monographs found in Chapter 6 of AACR are therefore generally applicable to MRDF insofar as the various points arise in this medium. The sections concerned, "Recording the title," provide guidance in derivation and transcription of the title from primary and other sources. However, one notable exception to the admonitions covering transcription of the title for monographs is here recommended: the provision requiring the use of brackets to enclose titles found other than on the title page should not be followed.

The fact that some MRDF are so inadequately identified within the file, certainly without a title page equivalent, is in part a result of the reason for their generation: they are the data of which individual research projects

are made, which lead to published results, summaries, interpretations -- they originate in fact as a means to an end. On reflection however, the contents of the files are realized to contain valuable data which it is felt advisable to preserve for some future use, a secondary consideration. The files become not just a means to an end but have themselves a raison d'etre -- but without a formal identification. Identification comes later in accompanying documentation, the published results, a report of the investigation or some such subsequent record, or concurrent announcement of the project. This lack of intrinsic identification characteristic of some MRDF would, according to Chapter 6, necessitate the constant use of brackets. It has been earlier noted where such identification may reasonably be found (paragraph 2 above and Appendix B). It is proposed now that titles found in such sources and used within the title area should not be enclosed in brackets.

(It may be noted that the Library Association Media Cataloguing Rules Committee has not required brackets on titles derived from accepted sources of information other than title page equivalents -- Non-book materials cataloguing rules, p. 19-20.)

Titles composed by the cataloger

The necessity of supplying a title identification for many MRDF seems unavoidable, particularly in the cataloging of files deposited with a library or archive upon completion of some research project but without documentation providing a title. Such titles composed by catalogers should be descriptive of the nature and scope of the contents of the work and stated in English as briefly as intelligibility permits. They should be enclosed in brackets to signal immediately their fabrication as in the cases of books and manuscripts, and should be documented in the note area of the catalog record as to the source of information upon which the invention is based. ("Title based on print-out." "Title derived from verbal description of depositor / compiler / or _____." Etc.)

(Here again it may be noted that Library Association Media Cataloguing Rules Committee has arrived at similar injunctions: Non-books materials cataloguing rules, p. 21.)

Acronyms

The frequent use of acronyms to identify MRDF in the literature which documents them requires consideration in a note of those acronyms which do not receive treatment as title proper, parallel or other title but which are considered essential to retrieval. So considered, they should be recorded in a note and traced.

Distinction is made here between an acronym and a data set name. The term data set itself has been shown to be of variable interpretation (Appendix C) according to different operating systems and is not applied consistently among the various systems. The make-up of the name itself is usually a result of local processing conventions and is often either whimsical or routine, and unrelated to the content of the file. The arbitrary composition

and assignment, local usage, possibility of change when physical form chan
(tape to disk, for example) or the possibility of several different data se
names being assigned to parts of a bibliographic unit, invalidate their va
as identifiers in a catalog record. As titles these data set names should
avoided in catalog records unless some documented evidence indicates the
intent of the originator to assign to the file a title which is coincident-
ally the data set name.

APPENDIX F

UNIFORM TITLE

The AACR defines a uniform title as "The particular title by which a work that has appeared under varying titles is to be identified for cataloging purposes." (p. 345) Weihs' Nonbook Materials repeats that definition word for word. (p. 88) Both authorities also concur in suggesting the use of uniform titles for "bringing together all catalog entries for a given work." (AACR, p. 145 and similar language in Weihs, p. 15)

In the interest of extending the area of common practice in cataloging the various media it should be possible to apply uniform titles to MRDF when they are works which would receive uniform titles in other media. As examples, the rules should permit use of uniform titles for

1.) a MRDF which is the same in terms of content as another work (i.e., another MRDF or another work in some other medium) which carries a uniform title;

2.) a MRDF which is the same in terms of content as other works carrying different titles;

3.) a MRDF which is a part of another work; and,

4.) a MRDF which is a translation of another work.

Where desirable, as in the case of MRDF which are parts or translations of other works, it should be permissible to add an appropriate explanatory qualification or designation to the uniform title in a manner similar to that suggested for monographs in the AACR. (Chapter 4 especially Rules 104 through 107).

Careful note should be taken of the permissive language used above. In the introductory notes to its chapter on uniform titles the AACR is explicit about the desirability of a flexible approach to the use of uniform titles.

> The need to apply the rules will vary from library to library and will depend on such factors as the general renown of the work; the number of editions, translations, etc., involved; whether or not entry is to be under title; whether or not the work was originally published in a foreign language; and the extent to which the collection is used for research purposes. (p. 145)

A similarly flexible approach should be taken when considering the use of uniform titles in relation to MRDF.

In at least two respects MRDF present special problems in the application of uniform titles.

First, in addition to containing translations of natural language contents, a MRDF may be a computer language translation of another MRDF. In such cases it should be permissible to append to the uniform title the computer language(s) of the translation.

EXAMPLE: ⌜LIBRARY SEARCH.⌝ (A program written in COBOL)

⌜LIBRARY SEARCH. PL/1⌝ (A copy written in PL/1)

Second, the informal manner of producing, distributing, and identifying MRDF at this point in time may exacerbate the problem of determining which of several possible titles should be selected as a uniform title for the file. MRDF frequently exist which are the same file in terms of content but have been given no title. Such MRDF are sometimes known by a variety of local titles which vary from location to location, user to user, etc. Caution should be employed in choosing a uniform title form among a group of informal titles. Determination of a uniform title should be postponed until the file in question is consistently cited by one title in the literature, or the author has designated a title, or formal distribution has resulted in a stable title. In other words, it is strongly recommended that some rational method be employed in determining which of several contending titles is to be designated as the uniform title.

EDITION

Because a Machine-Readable Data File (MRDF) is cataloged in such a manner that the catalog record may be filed into a multimedia catalog, the cataloging process parallels that for books in which it is customary to indicate the edition.[1] The purposes for attempting to apply the concept of edition to MRDF are to aid in uniquely identifying the work in hand and to aid in differentiating the presentation of the work in hand from all different presentations.

The concept of edition is often applied to books in terms of physical characteristics, because the differences in content (the underlying reason for making such distinctions) are often more easily described in physical terms. However in dealing with MRDF and the problems posed by the medium, the Subcommittee has excluded from the catalog record those elements describing physical characteristics comparable to the physical aspects of a printed volume (pagination, measurements of size, etc. -cf., Appendix B). By extension, these characteristics should then also be eliminated as determinants of edition for MRDF.

Thus, while the use of "edition" as an indicator of the physical characteristics of a MRDF is inappropriate, use of "edition" as an indication of the content characteristics of a MRDF is essential so that the user can know what content (i.e., edition) is being dealt with.[2] The edition of a MRDF changes only when the content is revised or augmented,

---

1. EDITION is variously defined and utilized in the description of printed material, but the most usual definition is "All the impressions of a work printed at any time or times from one setting of type, including those printed from stereotype or electrotype plates from that setting (provided, however, that there is no substantial change in or addition to the text, or no change in make-up, format, or character of the resulting book). One of the successive forms in which a literary text is issued either by the author or a subsequent edition." (ALA Glossary)

   The embodiment of a work in a particular typographical form. Different editions may embody an identical text, or varying texts. (IFLA per Librarians' glossary, 1971)

   "An edition consists of all the copies of a book printed from the same setting of type. A different edition may, or may not, involve revision of the text of the work." (Dunkin, p.50)

   By analogy, copies of a MRDF made from the same master file or from identical copies of it belong to the same edition.

2. For purposes of describing MRDF the Subcommittee has selected logical records and program statements as suitable surrogates for physical collation in the catalog record.

i.e., the structure of the content's representation is altered in such a way that a printout of the file is substantively different (see 3, 4 below). However, a rearrangement of the records in the file according to a different principle constitutes a different work.[3]

The primary source of information regarding the edition of a MRDF is the file itself, a display thereof, or reliable documentation (cf., Appendix BB). Otherwise, the information will be enclosed in brackets or recorded in a note according to accepted cataloging practice.

In addition to the general guidance provided in AACR, eight criteria (based on AACR practice and substituting (where necessary) logical records for collation) may be used to aid in differentiating editions of MRDF. When at least one criterion has been met, an edition is said to exist.

1. The presence of an edition statement (relating to content and/or structure) on the internal machine readable label or accompanying documentation.

   2d ed. --
   Rev. ed. --
   OSIRIS ed. --
   SPSS ed. --

2. The details of the content of the MRDF are different: additions, deletions, updating, revision, cleaning, and such other terms as customarily convey changes in content.

   "The new test reel [for 1970 census first count summary tapes] presents data from file A of the first count only...changes...have necessitated imputation of certain data."

3. The number (i.e., count) of logical records is different.

4. The size of the logical record is different.

---

3. Note: A corollary of edition is the concept of ISSUE. The issues of an edition are customarily identified by physical means (printing, binding, etc.) or by content (corrections, minor alterations, minor additions). For MRDF, issue is limited to corrections and to such alterations and additions that do not affect the structure or size of the file; such issues may be added to entry for the issue cataloged first, with variations specified.

5. The editor or compiler named is different.

6. The identity of the producer is different; however, the "producer" must represent more than a copier or distributor.

   The 1970 Census data is available in two MRDF forms, the so-called "Bureau format" and "compressed"; the imprints for these are:

   Washington, D.C. : U.S. Bureau of the Census, 1973.

   Washington, D.C. : U.S. Bureau of the Census, 1973 : Dist. by DUALabs, Rosslyn, Va.

   Rosslyn, Va. : DUALabs, 1973.

7. A different program language may indicate an edition or a new work. A literal translation from one program language to another would be an edition. In the absence of evidence to the contrary, treat as an edition.

8. A different text of a work was used as the basis for input. (Identify text used in a note.)

Note: The following are not criteria by which editions of MRDF may be differentiated.

1. The fact that the file in hand is a copy. (A note on the order of a Photocopy note may be used: "Copy. Los Angeles, Calif., UCLA Campus Computing Network, 1973.")

2. Copying dates.

3. Minor corrections (e.g., typographical errors, misspellings, incorrectly transcribed data). Such corrections would have no substantive effect on the file.

4. Physical characteristics.

5. A change in distributor.

References:

ALA Glossary. 1943.
Anglo-American Cataloging Rules. 1967.
Dunkin, Paul S. Cataloging U.S.A. 1969.
Librarians' Glossary. 1971.

PRODUCTION AND DISTRIBUTION

Because Machine-Readable Data Files (MRDF) usually do not pass through the customary writing, editing, and printing sequence enjoyed by books, it is better to describe MRDF in terms of Production and Distribution rather than Publication and Imprint. Production has to do with the exercise of or the immediate overall responsibility for the physical processes whereby a MRDF is brought into existence (as opposed to creation which is a function of authorship). Distribution has to do with the formal merchandising or formal dispersal of a MRDF to customers and other recipients. Sponsorship by a person, firm, etc. whose sole relation to a MRDF is the financing of the production and/or distribution thereof is excluded from consideration here.

In recognition of the frequent separation of the production and distribution functions, data relating to both should be recorded whenever the functions are performed by different agencies. It is possible for a MRDF to have a producer only or both a producer and a distributor. It is further possible that the identity of the producer may be unknown; in such an instance, the identity of the distributor may be known and should be carefully distinguished from the producer.

Instances of simultaneous production or distribution by two agents in two countries may be dealt with in either of two ways: 1.) If the title page or documentation makes it clear that a duality exists, both are recorded in the catalog record, but 2.) If the duality is evident only by juxtaposing two identical files with different production and/or distribution information, then an "Also issued ..." note is made.

The primary source of information for the Production and Distribution Area is the file itself, a printout or CRT display thereof, or reliable documentation (cf., Appendix BB). Otherwise, the information will be enclosed in brackets according to accepted cataloging practice.

Care must be taken to distinguish producers and distributors from other sources. When the source from which a MRDF is obtained is other than the producer/distributor, a note may indicate the source of the copy in hand. When an intermediate agency has altered a MRDF, it may become a new edition requiring a different imprint (cf., Appendix G, Edition). However, physical changes made in copying a MRDF (e.g., changes in medium, density, or machine code) would not require a different imprint nor any other mention in the catalog record.

Each entity named in the Production and Distribution Area should be identified as to its role in the production and/or distribution process. The statement of production and distribution may consist of multiple names if joint or subsidiary responsibility is involved and the additional names are essential for the identification of the work.

The special requirements for recording the Production and Distribution data are discussed below; for further guidance, Chapters six (revised) and twelve (revised) of AACR should be consulted. The sequence in which the Production and Distribution data are recorded is 1.) Place of Production, 2.) Name of Producer, 3.) Date of Production, 4.) Name of Distributor, and 5.) Place of Distribution. Note: The Date of Production, recorded immediately after the production data, is generally unrelated to the distribution data; recording the Date of Production after the distribution data (cf. AACR, chap. 6 rev. 136) would lead to confusion between the Date of Production and the date of distribution (which is not recorded).

Place of Production. The Place of Production represents the town or city in which is situated the producer's place of business or locale of operation. If only a probable place of production is known, it should be followed by a question mark. The absence of a known place should be indicated, e.g., "s.l." (sine loco).

Name of Producer. The Producer is that person or body responsible for the production of a MRDF (cf., paragraphs number one, two, and five above) and should be so recorded when known. A computer center, unless specifically identified on the file or in the accompanying documentation as serving the function of producer, is not recorded. The absence of a known producer should be indicated, e.g., "s.n." (sine nomine).

Date of Production. The Date of Production is the year when the MRDF was produced. The date is recorded when available from the file itself or from reliable documentation; otherwise, an approximate date is ascribed. However, if the copyright date is at variance, it should be recorded as an additional identifier of the file. In the absence of a production date, the copyright date, if any, may be substituted (but with the prefix "c").

The AACR Chapter six subsection on "Date" provides direction in recording problematical dates. If there is a choice of dates from sources of equal authority, use the earliest.

Often the materials accompanying a MRDF will be of assistance especially as they purport to relate to the file itself. The following sources of information may be of further assistance in ascribing a date. However, the relative order of usefulness is likely to vary considerably according to the content of the file.

- Date as derived from supplementary and/or related file(s)
- Date of publication of the contents of the MRDF which could be prior, simultaneous, or subsequent
- Time period during which data were collected
- Time period represented by data
- Date of distribution.

Since distribution copies are often made on demand rather than stockpiled, and since copying from one machine-readable medium to another is a physical process not admitted to the catalog record for MRDF, generally only the date of the original (i.e., first) production of an edition should be recorded. The date of distribution should not be confused with the date of the physical copying or the date of production of the representation of the file in hand. (Also, see Appendix G .(Edition) as regards copying related to edition.)

Name of Distributor. The Distributor is that person or body responsible for the distribution of the copy of a MRDF and is recorded "Dist. by ..." or a phrase taken from the file or accompanying documentation. The absence of a distributor is not indicated.

Place of Distribution. The Place of Distribution represents the town or city in which is situated the distributor's place of business or locale of operation. If the place of distribution is the same as the place of production, it is not repeated. The absence of a place of distribution is not indicated.

Note: The rationale for the structure of the Production and Distribution Area is based on AACR, mostly on the section on "Imprint" and the related sections on "Place of publication", "Publisher", and "Date" (all from the chapter on "Separately Published Monographs"). This is preferred over the abbreviated but not incompatible "General Rules for Entry and Descriptive Cataloguing" of Nonbook Materials. The Library Association's newly published Non-book Materials omits specific consideration of MRDF pending consideration of this Subcommittee's work; however, this paper is generally not incompatible with its rule number 7 (General Rules).

Examples:

place of production, producer, date

Baltimore, Md.: Statistical Operations Branch, Social Security Administration, 1971.

Columbus : Center for Human Resource Research, Ohio State University, 1974.

place of production, producer, date, distributor, place of distribution

Washington, D.C. : U.S. Bureau of the Census, 1973 : Dist. by DUALabs, Rosslyn, Va.

place of production, producer, 2d place of production, 2d producer, date

Philadelphia : Independent Productions, London, Unicorn Productions, 1974.

place of production, producer, date, distributor, place of distribution, 2d distributor, 2d place of distribution

Philadelphia : Independent Productions, 1974 : Dist. by Eagle Sales, New York, and Bulldog Industries, London.

place of production, production statement, date, distributor, place of distribution

New York : Produced by Comp-U-Media for Humanistic Studies Center, 1973 : Dist. by Datacasters, Los Angeles.

date, distributor, and place of distribution ONLY

[s.l. : s.n.] 1974 : Dist. by the Inter-University Consortium for Political Research, Ann Arbor, Mich.

date of a MRDF probably produced in 1969 and distributed in 1974

[1969?]

References:

Anglo-American Cataloging Rules. 1967.
Anglo-American Cataloging Rules; revised chapter six. 1973.
Anglo-American Cataloging Rules; revised chapter twelve. 1975.
International Federation of Library Associations. International Standard Bibliographic Description, Monographs. 1974.
Nonbook Materials: The organization of Integrated Collections. 1973.
Non-book Materials: Cataloguing Rules; integrated code of practice and draft revision of the AACR British text part III. 1973.

APPENDIX 1

SIZE OF FILE AREA

In the course of proposing rules for the cataloging of machine-readable data files in libraries, this committee has reached certain conclusions, expressed in previous working papers, which bear repetition.

The first concerns the kind of record we expect to create as a result of the rules being proposed. It is to be as nearly a standard library catalog entry as is possible so that it can be readily integrated into the general catalog of the institution concerned. It is to be essentially a record which tells the library user that the institution has a machine-readable data file, created or compiled by a particular person or corporate entity, identified by a particular title, characterized by certain subject categories and having certain bibliographic characteristics which can be described briefly and which can assist the potential user in deciding whether to investigate the material further.

In addition to indicating that a given program or a given body of data exists, the catalog record should also inform the user where he can find additional information about the file and assistance in accessing it. It is assumed that every library will have a service point supplemental to the catalog where description of the attributes (documentation such as data abstracts, codebooks and records of physical and logical characteristics) of the file would be maintained and where arrangements could be made to use the file. The assumption of such a system of documentation is implicit in all of the discussion which follows. It stems from the nature of machine-readable data files which precludes visual examination for contents and relevance and which, in most cases, requires detailed descriptions of both file organization and mode of storage in order to facilitate meaningful use. A complete description of the physical characteristics of a file would exceed the normal limits of a catalog record, and would, in any case, not be relevant at first approach. Even were one to opt for such a lengthy record, the physical form of a machine-readable data file can easily be changed without affecting bibliographic identification. Catalog records cannot be reactive to frequent changes and thus would frequently misrepresent the file.

The committee has consistently recommended that highly variable and extremely detailed aspects of data file description be omitted from catalog records in favor of information indicating the intellectual content of the file, its approximate size and its availability. A decision was made, therefore, that in the area of the catalog record normally reserved for description of the physical aspects of the item in hand, only a designation of the relative extent of the content be entered rather than the physical details necessary for actual use of the file. This area is usually known as the Collation but for purposes of MRDF is call Size of File Area.

One of the alternatives which the committee considered for entry in the Size of File Area was a description of the file's contents in terms such as "39 states with 22 economic variables" or "a cross-section of 3,120 New Yorkers." Although such a statement might be useful to the catalog user, the difficulty of writing rules for this type of collation which would provide uniform and unambiguous results proved insurmountable. This information would more appropriately be included in the note area when necessary.

The committee, therefore, recommends an alternative which does not present this difficulty, namely, the use of "logical records"[1] as a useful and consistent indicator of the extent of the content of the file and of the amount of data to be handled by the user. The total number of logical records would be entered even in those cases in which the file is stored on more than one physical volume since the number of logical records is a permanent attribute of the file but the number of volumes is subject to change.

---

1 The definitions quoted below emphasize content as distinguished from physical attributes in describing logical records.

A collection of items independent of their physical environment. Portions of the same logical record may be located in different physical records. (ANSI)

A record from the standpoint of its content, function, and use rather than its physical attributes; that is, one that is defined in terms of the information it contains. IBM Data Processing Glossary, 1972.

A record whose scope, direction, or length is governed by the specific nature of the information or data which it contains instead of by some feature or limitation of the storage device that holds it. Sippl, Computer Dictionary and Handbook, 1966.

A record identified from the standpoint of its content, function and use rather than its physical attributes. It is meaningful with respect to the information it contains. Contrasted with physical record. Meek, Glossary of Computing Terminology, 1972.

The committee identified several advantages related to this recommendation: 1) since the term logical records has a consistent meaning regardless of storage medium, it permits an objective measure of file size which will not vary from catalog entry to catalog entry; 2) logical records represent a stable element on the catalog record because they are not affected by changes in block size, density, etc., of the file; 3) the number of logical records contained in a file, or at least an adequate estimate, frequently appears in the accompanying documentation or may be calculated from other data presented therein; in addition, it is possible to derive the number of logical records by computer; and 4) the term logical records is usually familiar to those who utilize machine-readable materials and, even to the uninitiated, its meaning should be more or less clear within the context of the full catalog entry.

Any additional descriptive information will be relegated to the note area, the nature of which will depend on the particular machine-readable data file being cataloged. Typically, the note will define the logical record, especially in the case of text files; explain variability in the length of logical records and record details concerning the incomplete or otherwise imperfect condition of the file.

With regard to obtaining information concerning the size of the file, the machine-readable data file and accompanying documentation are the preferred sources, although other reliable descriptions, whether published or not, may serve this purpose. Whenever file size has been estimated or determined from sources other than the file or its accompanying documentation the cataloger should include a note that the size of the file has not been verified. If an estimate is provided, this estimate would appear in Size of File Area with "Size of File not verified" entered in the note area. In cases of very large collections which are received without information concerning the number of logical records, one could say "received as 100 tapes" (cf. MRDF 47); in this instance, Size of File Area would be blank.

An exception is made to the use of the term logical records in the case of source programs which are customarily described in terms of the number of statements they contain in the source language.[2] Program statements are, in effect, the equivalent of logical records. Since the number of statements in a program is meaningless without a statement of the programming language used and since this information can be stated concisely, it is recommended that the programming language, if known, be stated in the Size of File Area. For example, 500 ANSI COBOL statements or 400 FORTRAN IV statements with Assembler Subroutines. A more complex situation would be explained in a note.

---

2 Source language: The original form in which a program is prepared prior to processing by the machine. It is usually referring to a program written in an advanced programming language as opposed to machine language coding. Meek, Glossary of Computer Technology.

Since the size of an object program[3] cannot usefully be expressed in terms either of logical records or program statements, the Size of File Area would be omitted in these cases. Instead a note would identify the file as an object program and specify the computer on which it runs e.g., "object program (IBM360/40)."

A second exception occurs in the case of multi-file works which are bibliographically related and cataloged as a single entity. These would include, e.g., a multi-nation study with several data files or a collection of short stories each stored as a separate file. If the entire work can be adequately described with a simple expansion in the Size of File Area such as, "six files each with 500 logical records" or "3 files (250, 8500, 9200 logical records), this should be done. In more complex situations, the Size of File Area should contain the number of files and the expansion should appear in a Contents or Summary Note.

The second element of the Size of File Area is a description of material accompanying the work being cataloged, intended to be used in conjunction with it. Such materials may include computer programs, text, data, codebooks, data abstracts, memoranda, program listing, project descriptions, etc., i.e., any material designed for use with the file(s) being cataloged. When such materials can be used alone or with more than one bibliographically independent MRDF they would be cataloged separately.

Accompanying material in machine-readable form would best be described in terms of a word or phrase, following the description of the work being cataloged and preceded by an ampersand indicating the nature of the material and a statement of the size of the file given in parentheses after the characterizing term. All of the rationale for providing a measure of the file size for the item being cataloged exists for the item considered accompanying material. For example:

1) A data file accompanied by a program file:
   30,000 logical records and program (300 COBOL statements)

2) A program accompanied by a data file:
   300 COBOL statements and data (30,000 logical records)

3) A program accompanied by a multi-file work:
   400 COBOL statements and data (3 files (250, 8500, 9200 logical records))

The use of such terms as logical record, program, COBOL statements, should indicate to the catalog user that the accompanying material is in machine-readable form.

---

3 Object program: The computer language program prepared by an assembler or a compiler after acting on a programmer written source program. Jordain, Condensed Computer Encyclopedia, 1969.

Accompanying material not in machine-readable form should be treated in the same fashion but instead of the number of logical records the catalog would indicate the extent in the manner appropriate to the medium. The term "associated documentation" would be used to describe an assemblage of descriptive, narrative, etc., material which might normally accompany the file with the major item, e.g., t codebook distinguished; or, if the accompanying material is simply that " & codebook (xxx p.)". Should further elaboration seem necessary, a brief note should be composed to cover the material. Should this material be of minor importance and of a nature typical of such accompaniments, the simple "& associated documentation" will do. For example:

1) A data file accompanied by a printed codebook:
   30,000 logical records and codebook (240 p.)

2) A data file accompanied by both a printed and a machine-readable codebook
   30,000 logical records and codebook (240 p.), codebook (990 logical record

3) A program file accompanied by a manual:
   5,000 FORTRAN IV statements and manual (100 p.)

4) A multi-file work with an assortment of documentation:
   22 files and associated documentation

The committee believes that the use of the term logical records as a means of indicating the size of the item being cataloged and, the accompanying materials, would avoid complicated rules and yet would allow for the variable content of data files to be effectively described. It provides useful information to the user and relieves the cataloger of the responsibility of providing a detailed collation especially when information is lacking or unclear. This alternative presents the catalog user with a satisfactory picture of the extent of the file's contents, assists him in identifying the work and tells him something of its nature. Further, it should aid him in differentiating among various editions and ensure that all those parts of the work which belong together are so recorded.

ALTERNATE RULE FOR ARCHIVAL MATERIALS, EXEMPLARS, AND OTHER EXCEPTIONAL MRDF

The Subcommittee has recommended in Appendix B that the physical characteristics of MRDF not be described in the catalog record. This is the most practical alternative for handling the vast majority of MRDF. However, in certain instances, a physical description may be necessary.

Archival files, especially in records management programs and "rare book" type collections may need to be described in physical terms as well as content. For example, it is customary in records management programs to retain only a percentage sample (in varying amounts but typically a small percent). For such exceptional MRDF (e.g., Burroughs ledger sheets with a magnetic tape strip on the back or plastic credit cards with a similar magnetic tape strip) an entire file would probably not be retained, only a few exemplars.

For these archival materials, exemplars, and other exceptional MRDF, an alternate physical description to fulfill this special need is recognized. However, this should apply only to a select few MRDF which are demonstrably exception and, most important, physically stable.

Examples:

For a truly voluminous file for which obtaining a measure of the number of logical records is not feasible, leave the size of file area blank and use a descriptive note:

"Received as 2,000 reels of magnetic tape."

"8,300 boxes of punched cards."

For exemplars which are retained for their physical characteristics, leave the size of file area blank and use a descriptive note:

"Sample collection of 100 ledger sheets." (Note: The type of ledger sheets is described in the body of the catalog record.)

"75 plastic credit cards with magnetic strip on back."

NOTES

A detailed review of the provisions for notes in AACR does not show any conflict with the proposals being made by the subcommittee with regard to MRDF. While many of the conditions when notes are indicated by AACR are not possible with MRDF, in no case is a note prohibited by AACR when it is recommended by the subcommittee.

Most types of notes specified in revised Chapter 6 AACR could apply either directly, or by some extension, to MRDF. The major exceptions center around notes concerned with the physical characteristics of books as such. The provisions of Rules 144, 145, 147 and 150 would appear to be irrelevant to the cataloguing of MRDF. Similarly, with reference to serials, Rules 169 and 171 would not apply.

The provisions of Rule 143D2 dealing with translations and with variations of title in different editions might be required in cataloguing a MRDF which comprised literary text or which had received international distribution.

Whereas MRDF do not yet bear any kind of international standard number, codebooks associated with them sometimes do; Rule 149 would allow the codebook ISBN to be recorded. Any other standard number should also be recorded.

The thesis note, as such, would not seem to be useful, but with a more inclusive wording for Rule 146B2, it might be valid where a MRDF accounted for a substantial part of the thesis work.

Rule 148, revised Chapter 6, AACR, states:

> 1. Either all of the contents or a part of them are specified in the catalog entry if it is necessary to bring out important parts of the work not mentioned in the title and statement of authorship area, or to give a fuller and more detailed description of the contents than the title supplies...
>
> 2. Contents are especially necessary for works in several volumes...

These provisions would appear quite broad enough to cover any formal or informal contents notes which might be required for a MRDF. Contents notes would seem to be particularly valuable in multi-file works, which are in many ways analagous to multi-volume works.

Rule 229.2M, revised Chapter 12, AACR, states, in part:

> The contents note is particularly appropriate if the titles of the parts are sufficiently descriptive to serve also as a summary. Otherwise, prefer a summary to a contents note; both are not normally recorded. Formulate a contents note according to the general rules (see 148). Include after the title of each part any data normally belonging in the other notes or in other areas of description, if a more concise or clearer presentation of information can be achieved in this way.

This statement clarifying the conditions under which a Contents note is to be preferred over a Summary note (cf. MRDF-17) would seem equally useful in cataloguing MRDF. Further, the first example after the passage quoted above shows collation information following each title in a formal contents note. Where individual files in a multi-file work have clear titles of their own, it would be useful to extend this approach to MRDF contents notes.

A check of Weihs' Nonbook materials did not uncover any useful notes which had not been considered in either AACR or in subcommittee discussion. A similar review of the draft AECT Standards (App. L) identified two areas where there is a divergence from the subcommittee approach.

1. (p.4)

"C. Date. The original release date is given. An approximate date may often have to be supplied. Sources from which the approximation is derived such as dates of supplementary files and time periods of data collection, are given in a note."

This seems to make the provision of a date mandatory and would probably lead to more notes dealing with date than the pattern projected in App. H.

2. (p.5)

"Notes
...B. Accompanying and/or descriptive material. The existence of a separate code book or other supplementary documentation provided with the file is noted. Such accompanying material is often the source for physical and content information. The notation may be a simple statement of the purpose of the accompanying material (e.g., With format; With code book) or may include identifying bibliographic details and pagination.

This departs considerably from our approach to accompanying material, which places it in the MRDF Size of File area (following the ISBD area pattern) and only calls for a note when the accompanying material becomes too complex to describe succinctly.

Our recommendations are completely compatible with the revision of Chapter 12, AACR. In particular, Rule 229.2B provides for a note when the title is taken from a source other than the work itself or material accompanying it. With a MRDF this might well be the catalog of the holdings of a specific data archive. Rule 229.2D, Extension of physical description, meets MRDF requirements arising from complex information beyond the usual capacity of the size of file area. Rule 229.2E, Accompanying material, is very close indeed to the position the subcommittee has arrived at on this information, much closer than that in the AECT Standards mentioned above. Rule 229.2F, Related works, is compatible with our approach to bibliographic relationships. Rule 229.2L is equivalent to the Summary note we are recommending for virtually all MRDF. Rules 229.2G-229.2K, which are mainly intended for 'motion pictures or similar works' seem irrelevant to MRDF.

A check of Non-book materials cataloguing rules produced by the Library Association Media Cataloguing Rules Committee showed that MRDF had been deliberately excluded from the work. As with the revision of AACR, Chapter 12, the general rule for notes, Rule 10, allows a broad range covering "any information required to explain the nature, scope and relationships of items that is not shown in the formal description, and any other information useful to the catalog user." Among the note categories specifically mentioned are summary, contents, related materials, edition and imprint. In addition, there are specific rules for notes for graphics (GT10), motion pictures (MP10) and sound recordings (SR10). As is to be expected, most of these are so closely related to the medium concerned that they are irrelevant to MRDF. Analagous situations may occur, but the conditions described, as such, do not apply.

The LA/MCR Draft rules: Computer media (MRDF-42) gives only a sketchy indication of notes to cover: content description, sources of additional information, accessory materials needed for file utilisation, related files, minimum core storage for utilisation, and copyright restrictions on use. All these can be accommodated in the framework projected by this subcommittee, except the note on core storage requirements, which we have tended to think belonged in another level of documentation, more technical and more detailed than the standard catalog record. Furthermore, it seems reasonable to assume that a MRDF recorded in an institution's multi-media catalog will be usable in the computing facility of that same institution.

A great many notes would appear to be possible, and frequently necessary, in the cataloguing of MRDF, because of the lack of a reliable title, because of the ease with which a file can be changed to produce a new edition of that file, and because of the frequent need to relate a MRDF to its accompanying documentation. Amongst the notes explicitly or implicitly sanctioned by AACR are the following which may prove to be particularly applicable, together with these additional indicated notes.

1. Title and statement of authorship area

   A. The primary source of the title will always require a note, except where the title is taken from the user header label, or from a printout of the user header label.

      Title taken from accompanying codebook.

      Title derived from verbal description given by compiler.

      Title based on printout of file.

   B. Titles commonly associated with a MRDF, but not selected as the title proper or other title, including acronyms, but excluding data set name(s), would be recorded in a note.

      Also known as: NLS, and Parnes study.

      Title on code book: New Democratic Party of Ontario, 1967.
      (Title proper: Ontario NDP support study)

2. Edition area

   A. Content changes made in a MRDF that would reflect changes in a printout should be indicated in a note.

      Mnemonic tags substituted for numeric tags.

      Transactions of individual respondents aggregated by group for each stock.

      Raw data edited by Institue staff, using error detection programs.

      Formatted in 35-line, 60 characters per line, page images.

      Capcen compacted file of 1970 census data.

   B. When a specific edition of a literary work has been used for input, this should be identified in a note.

      Based on C. Valerii Catulli Carmina, ed. by R.A.B. Myners, Oxford Classical Text, 1960.

C. Program version and/or level.

SPSS 5.2

FORTRAN IV, level H.

3. Production and release area

A. If the sponsor, commissioning agent, programmer or systems analyst has not been included in the statement of authorship, or in the production and release area, and needs to be mentioned at all, this should be in a note.

Prepared for the Manpower Administration, U.S. Dept. of Labor under contract no. ...

Simulation revised and reprogrammed in BASIC by John Smith for use in an on-line time-sharing environment.

B. Dates when copying was performed by the producer or distributor need not be recorded. Date of copying of files from other sources may be recorded in a note when the provenance suggests that file content may vary from the original.

Data for 1945-60 from Duke University copy.

Copied on-line from State University Computer Center, June 1975.

1969 edition corrected 1975 using update program received 1972.

C. The source of the date in the production and release area may be provided in a note at the discretion of the cataloguer.

Production date taken from correspondence with the principal investigator.

4. Size of file area

While it may often be most convenient to include information about the size of file area in the Summary note, or as part of the Contents note, it may also be an independent note.

A. Extent of file note.

Size of file not verified.

Includes all publicly available time series from the ... data base.

Each file contains approximately 2000 logical records. [Multi-file work].

File size varies between 3000 and 4000 logical records. [Multi-file work].

B. Object program note.

Object program (IBM 360/40)

C. A Note may be needed to indicate that records vary in nature within the file. The fact that records are of fixed length or of variable length need not be noted.

Trip records include travel data and demographic data; records from 'no-trip' households include demographic data only.

D. A note may provide descriptive information too extensive to be put into the size of file area.

Data accompanied by a series of 5 programs in PL/I, with Assembler subroutines.

Codebook available in both English and French.

E. In those rare cases where an estimate of the number of logical records in a MRDF cannot be made, and the cost of a computer count is prohibitive, it may occasionally be necessary to leave the size of file area blank and record available information in a note.

Received as 2,000 reels of magnetic tape, 800 bpi.

8,300 boxes of punched cards.

F. For exemplars which are retained for their physical characteristics, leave the size of file area blank and use a descriptive note.

Sample collection of 100 ledger sheets.

75 plastic credit cards with magnetic strip on back.

5. Bibliographic relationships

A. A note may indicate that a file is derived from another file, or from a publication.

Created from files used to prepare tables for Census of housing, V.3, Block statistics.

Data derived from the U.N. Statistical yearbook.

Source of data is Human Relations Area File.

Data previously issued in printed form under title .../.

Digital representation of musical score.

B. If the documentation indicates there is a parallel publication in another form, bearing a different title, this may be noted.

Microform edition has title: ...

6. Restrictions on access

If the material is not generally available, a note should indicate the nature of the restrictions.

Available only to faculty, staff, and students of Blank University.

Available only with the permission of Dr. X.Y. Smith.

File closed until January 1979.

7. Contents note

EXAMPLE:

CONTENTS: file 1. Idaho (985 logical records). --- file 2. Montana (1102 logical records). --- file 3. Oregon (1158 logical records). --- file 4. Washington (2544 logical records).

SUMMARY

Summary notes have been used primarily in audio-visual cataloging to aid the library patron on selecting items which cannot be "readily examined or adequately described by title and/or series statement" (Weihs, et al..., p. 14). Non-book Materials Cataloging Rules (p. 51) in referring to motion picture notes indicates that a summary note is necessary and "should describe concisely and objectively the content of the film, and be specific enough to serve as the basis of classification." The general rules in the same document (p. 28) would include audience level in the summary note, regardless of medium. This does not seem to be a requirement which applies to MRDF in general. The AACR rule 229.2L (rev. Chapter 12) also gives comparable descriptions of summary note contents.

Since MRDF are difficult to examine, the summary should assist the user in determining the usefulness of the data file for his/her needs. A summary note should be considered necessary only if the title and/or other notes provided on the catalog card do not make clear to the user the content of the file. If the titles of individual parts are sufficiently descriptive to serve as a summary, then a contents note would be given instead (cf. Appendix K); normally both contents and summary notes are not given.

As with other media, the note should be in paragraph form and begin SUMMARY:. The note should include a description of the contents given succinctly and objectively. For example:

> SUMMARY: Information on higher civil servants in the U. S. federal agencies, including personal characteristics, educational background, occupational mobility.

For programs, the purpose of the program would form the basis of the summary note.

> SUMMARY: Utilizing a random number generator, this program simulates a crap game with the following contraints: a maximum of thirty throws or loss of money by the player.

Content description should be limited to the scope of the MRDF and should not include data description applicable only to the total research project which compiled the MRDF. However, it should indicate the relationship of the MRDF to that total project. E. g.,

> SUMMARY: Responses of New York City adults to Harris study questionnaire [no.] 1925 used during April and May, 1969.
>
> TITLE: Sources of racial and religious tension in New York City

MRDF may also require expansion of the size of file area which may fit more

logically into the Summary than in an independent note. However, such information should not duplicate an independent note nor be used solely to replace a suitable independent note (cf. Appendix K).

SUMMARY: Each record represents selected fields from only those records less than 2049 characters issued on LC MARC tape vol. 6, no. 5.

The organization of the file may be indicated in describing the content of the file within the Summary.

SUMMARY: Economic data arranged by nation-years 1970 to 1975.

SUMMARY: Monthly records for years 1900-50 for 23 nations.

The Summary may, therefore, amplify intellectual content and/or file organization which cannot be better explicated elsewhere in the catalog entry.

Glossary

This glossary identifies the terms most commonly used in cataloging machine-readable data files (MRDF) which the Subcommittee* felt would be needed by those reading this report. Definitions of these terms are essential to the understanding of the Subcommittee's report and any cataloging rules that would be based on it.

The Subcommittee recommends that the scope of the Anglo-American Cataloging Rules (AACR) glossary be expanded to permit the inclusion of terms relating to machine-readable data files as well as terms of a similar level of detail for the other media. The glossary should contain all terms needed by catalogers when applying AACR.

The Subcommittee also recommends that the Anglo-American Cataloging Rules rule and page references be given for each term in the glossary. Each reference should indicate the rule and the page number where the term first occurrs and/or is discussed in detail.

In the glossary that follows, a brief definition for each term is provided to convey its meaning in the context in which the Subcommittee has used the term and/or to identify the term for catalogers approaching data processing for the first time. The Subcommittee recognizes the limitations of this treatment which does not pretend to define terms conclusively. Glossaries and dictionaries consulted by the Subcommittee (see bibliography at end) are not in agreement themselves and usually provide detail and syntax calculated to confuse the inexperienced consultant. Rather than recommend a particular publication, the Subcommittee does commend to the cataloger and reputable recent work, should the complexities of sorting out terms in the documentation become overwhelming. In most cases, it is hoped that this glossary will be of sufficient assistance.

* The Subcommittee wishes to acknowledge the assistance of Peter Watson, Data Services Coordinator, UCLA Research Library, in compiling the glossary.

ALGOL	Acronym for Algorithmic Oriented Language or Algorithmic Language, a scientifically- and mathematically-oriented programming language.
Alphameric	See Alphanumeric
Alphanumeric	A contraction of alphabetic-numeric indicating a set of characters which may include letters, numerals and/or special symbols.
ANSI	Abbreviation for American National Standards Institute, a non-regulatory body which develops and publishes standards for the United States.
APL	Abbreviation for A Programming Language, a mathematically-oriented programming language used in on-line inter-active systems.
ASCII	Abbreviation for American Standard Code for Information Interchange, a standard set of bit patterns used to represent symbols, characters, and digits for information interchange among data processing systems, communication systems, and associated equipment.
Assembler	A computer program which converts symbolic instructions into machine language instructions on a one-to-one basis.
Assembly language	A machine-oriented programming language which allows the program to express each computer instruction in symbolic form.
BCD	See Binary Coded Decimal
*Binary	A characteristic, property, or condition in which there are only two possible alternatives. Also pertains to a numbering system which uses the base of 2 and only 2 digits, 0 and 1.
*Binary Coded Decimal	Abbreviated as BCD. A machine notation in which each of the decimal digits is represented by a unique binary code.
Bit	1. A contraction of binary digit, a single digit in the binary number system. 2. The smallest unit of information capacity of a storage device.
Bits per inch	Abbreviated as bpi. Usually synonymous with characters per inch of a magnetic storage device. A measure of bit density.
Block	A quantity of data such as words, characters or bits, treated as a physical record for computer processing purposes. May be smaller than, equal to, or greater than a logical record.
Bpi	See Bits per inch
Byte	A group of adjacent binary digits usually representing a single alphanumeric character.

* Recommended for inclusion in the Glossary of the Anglo-American Cataloging Rules

Cartridge	A permanently encased reel of film or magnetic tape which has the ends joined together to form a loop; used for filmstrips, motion pictures, sound recordings, videorecordings, and machine-readable data files (Adapted from AACR, Chapter 12, rev., p. 49)
Cassette	A permanently encased film or magnetic tape that runs reel-to-reel; used for motion pictures, sound recordings, videorecordings, and machine-readable data files (Adapted from AACR, Chapter 12, rev., p. 49)
Cathode ray tube	Abbreviated as CRT. In data processing, a computer display device similar to a television set.
Central processor	See Central processing unit
Central processing unit	In data processing, the part of the computer system which contains the circuits that control the computer and execute program instructions.
Channel	In data processing, the number of bits of information across a magnetic tape. Sometimes called track.
Character	In data processing, a letter of the alphabet, a numeral, a punctuation mark or other symbol.
Characters per inch	Abbreviated as cpi. The number of characters which may be stored along an inch of a magnetic storage device. See also Bits per inch.
CIM	Acronym for Computer Input Microfilming, a process in which microfilmed information is input directly into the computer using character recognition.
COBOL	An acronym for Common Business-Oriented Language, a problem-oriented programming language.
CODASYL	An acronym for Conference on Data Systems Languages, a committee organized and sponsored by the United States Department of Defense.
*Codebook	A manual which describes the organization and content of one or more data files.
COM	Acronym for Computer Output Microfilm, (ing, or) a process in which output from the computer is directly produced on microfilm without the intermediate print out and photographic copying steps.
Compiler	In data processing, a computer program which converts each program statement of the source program into many machine language instructions.
*Computer	A device capable of processing information by executing arithmetic and logical operations at high speeds.

* Recommended for inclusion in the Glossary of the Anglo-American Cataloging Rules

Computer program	A set of instructions or statements which enable a computer to perform a given task.
Computer-readable data file	See Machine-readable data file
Computer tape	See Magnetic tape
Core	In data processing, commonly used to denote the internal high speed storage of a computer in which programs and data are stored for access by the central processing unit.
Cpi	See Characters per inch
CPU	See Central processing unit
Data	The quantities, characters, or symbols on which operations are performed by computers and other automatic equipment, and which may be stored or transmitted in the form of electrical signals, records on magnetic tape or punched cards, etc.
Data abstract	A brief summary of the information contained in the codebook or other documentation of a machine-readable data file.
Data archive	An organization which produces, stores and/or distributes machine-readable data files.
Data base	A large file or group of files containing information appropriate to a variety of applications.
Data set name	In certain computer systems, the name the software recognizes as the identifier of a given machine-readable data file.
*Date of release	See Release date
Density	In data processing, the number of characters contained per unit length, such as 1600 characters per inch on a magnetic tape.
Direct access storage device	A medium of information storage in which the access time is independent of the location of the information on the device, as distinguished from sequential access.
Disc	See Disk
Disk	In data processing, a direct access storage device. Synonymous with disc.
Disk pack	In data processing, a set of disks mounted on a common spindle.

* Recommended for inclusion in the Glossary of the Anglo-American Cataloging Rules

*Distributor — In data processing, a person or corporate body primarily responsible for disseminating copies of a machine-readable data file.

*Documentation — In data processing, collection of documents or information describing a particular program or file(s).

EBCDIC — Abbreviation for Extended Binary Coded Decimal Interchange Code which is used to represent information in machine-readable form.

*Field — One or more bits or characters constituting a record subdivision reserved for data of a particular type.

*File — An organized collection of records that are related in some way and treated as a unit.

*File size — See Size of file

*File structure — The way in which a particular file is organized, e.g. alphabetical, hierarchical, etc.

*Fixed field — A field in which the length of the information is the same for each occurrence of the field in all records in a machine-readable data file. Contrasted with variable field.

*Fixed length — A characteristic of records in a machine-readable data file signifying that the size of each record is constant for the file. Contrasted with variable length.

*Format — A predetermined order or arrangement of data in a record.

FORTRAN — An acronym of Formula Translation, Formula Translating System or Formula Translator, a scientifically - and mathematically oriented programming language.

Hard copy — A document, in a form suitable for human beings to read without the aid of a machine. Usually refers to paper printout from a computer.

Hardware — In data processing, the physical components of a computer. Contrasted with software.

Header label — See Label

Hexadecimal — A numbering system which uses the base of 16 and the digits 0 to 9 and A,B,C,D,E,F.

Input — 1. The data to be entered into a computer for processing.
2. The process of or device for entering data into a computer.

*Instruction — In data processing, one of a series of computer program steps which specifies a function or operation to be performed.

* Recommended for inclusion in the Glossary of the Anglo-American Cataloging Rules

K.	An abbreviation of Kilo-, 1000. In data processing, frequently a measure of computer storage or file size (kilobyte, kilobit, etc.).
*Label	A machine-readable identifier which frequently provides information about the contents of a volume or a file.
*Logical record	A group of words, characters, or bits identified as a unit on the basis of content function and use rather than physical attributes. To be distinguished from physical record.
M.	An abbreviation of mega-, 1,000,000. In data processing, frequently a measure of computer storage or file size (megabyte, megabit, etc.).
*Machine language	That system of internal coding which can directly cause a computer to operate or perform a function.
*Machine-readable data file	Body of information coded by methods that require the use of a machine (typically but not always a computer) for processing Examples include files stored on magnetic tape, punched cards, with or without a magnetic tape strip, aperture cards, punched paper tapes, disk packs, mark sensed cards, optical character recognition font documents, etc.
Machine-readable data set	See Machine-readable data file
Magnetic disk	See Disk
Magnetic-ink character recognition	Abbreviated as MICR, the process in which characters printed in magnetic ink are recognizable by a machine. Uses a special font.
*Magnetic tape	In data processing, a continous magnetized material used to record data.
Mark sensing	The automatic recognition of a mark made with some electrically conductive material on pre-determined positions of a punched card or sheet of paper.
*Memory	In data processing, the portion of a computer where information is stored and retrieved.
MICR	See Magnetic ink character recognition
*Multi-file work	A bibliographic entity which consists of more than one machine-readable data file.
Natural language	A language whose rules are based on current usage rather than on prescribed usage, e.g. English, French, etc. rather than FORTRAN, ALGOL, etc.
*Object program	The computer language program prepared by an assembler or a compiler after acting on a programmer-written source program (Jordain, p. 348).

* Recommended for inclusion in the Glossary of the Anglo-American Cataloging Rules

OCR	See Optical character recognition
Octal	A numbering system which uses a base of eight and the digits 0-7.
*Off-line	In a computer system, not being under the direct control of the central processing unit.
*On-line	In a computer system, being under the direct control of the central processing unit.
Optical character recognition	Abbreviated as OCR, the identification of printed data by means of a light sensing device.
Output	1. The data that have been processed by a computer. 2. The process of receiving information from a computer. 3. The device for receiving information from a computer.
*Paper tape	A medium for information storage in which data are represented by a pattern of holes punched on a roll of paper tape.
*Physical record	A collection of data defined in terms of physical parameters, rather than logical content. Contrasted with logical record.
PL/1	Acronym for Programming Language-1, a sophisticated business- and scientifically- oriented programming language.
*Printout	A form of output in hard copy from a computer.
*Producer	The person or corporate body having the immediate overall responsibility for the physical process whereby a machine-readable data file is brought into existence.
*Programming language	A language in which computer programs are written.
*Program	See Computer program
*Programmer	A person responsible for designing, writing, testing, and maintaining computer programs.
*Punched card	A medium for information storage in which data are represented by a pattern of holes punched in a card.
Punched paper tape	See Paper tape
Punched tape	See Paper tape
*Read	In data processing, to obtain data from one medium of storage and interpret and/or transfer it to another medium.
Record	See Logical record, Physical record

* Recommended for inclusion in the Glossary of the Anglo-American Cataloging Rules

*Record length	The size of a record, usually given in units such as words or characters.
*Release date	The time at which a machine-readable data file was made available to the general public.
*Size of file	A designation of the extent of the content of a machine-readable data file.
*Software	In data processing, a term referring to a computer program or system of many computer programs. Contrasted with hardware.
*Statement	In computer programming, a meaningful expression or generalized instruction in a program.
*Source program	A program written in a language such that it required assembly or compilation before a computer can execute it.
*Tape	See Magnetic tape, Paper tape
Track	The path on a magnetic storage medium on which data is recorded. See also Channel.
*Variable field	A field in which the length of the information may vary for each occurrence of the field in the records of the machine-readable data file. Contrasted with fixed field.
*Variable length	A characteristic of records in a machine-readable data file signifying that the size of each record may vary throughout the file. Contrasted with fixed length.
Volume	In data processing, a physical unit of external storage such as a disk, a reel of magnetic tape.
Word	In data processing, an internal processing unit unique to a specific computer. May vary in length from one to many characters.
Write	In data processing, to record data from any source onto an external storage medium such as a magnetic tape or disk.

* Recommended for inclusion in the Glossary of the Anglo-American Cataloging Rules

The following entries represent works cited in the various appendices. For a list of dictionaries used to compile the Glossary, consult that section of the report.

BIBLIOGRAPHY

American Library Association. Editorial Committee. Subcommittee on Library Terminology. A. L. A. Glossary of Library Terms. Chicago, American Library Association, 1943. 159 p.

American National Standards Institute. American National Standard Vocabulary for Information Processing. Rev. ed. New York, 1970. 128 p.

Anglo-American Cataloging Rules. North American text. Chicago, American Library Association, 1967. 400 p.

Anglo-American Cataloging Rules: chapter 6 (rev.), Separately Published Monographs. North American text. Chicago, American Library Association, 1974. 122 p.

Anglo-American Cataloging Rules: chapter 12 (rev.), Audiovisual Media and Special Instructional Materials. North American text. Chicago, American Library Association, 1975. 56 p.

Association for Educational Communications and Technology. Information Science Committee. Standards for Cataloging Nonprint Materials: Machine Readable Data File. draft, 4th ed. 1974. 12 p.

Burchfield, R. W., comp. A Supplement to the Oxford English Dictionary. Oxford, Clarendon Press, 1972- v.

Clason, W. E. Elsevier's Dictionary of Computers, Automatic Control and Data Processing in 6 languages... Amsterdam, New York, Elsevier Scientific Publishing Co., 1971. 474 p.

Dunkin, Paul Shaner. Cataloging U. S. A. Chicago, American Library Association, 1969. 159 p.

Dunkin, Paul Shaner. How to Catalog a Rare Book. Chicago, American Library Association, 1951. 85 p.

Dunkin, Paul Shaner. How to Catalog a Rare Book. 2d ed. Chicago, American Library Association, 1973. 105 p.

International Business Machine Corporation. Data Processing Division. Data Processing Glossary. 5th ed. White Plains, N. Y., 1972. 144 p.

International Federation of Library Associations. Joint Working Group on the ISBD(S). ISBD(S): International Standard Bibliographic Description for Serials. London, IFLA Committee on Cataloguing, 1974. 36 p.

Jordain, Philip B. and Michael Breslau. Condensed Computer Encyclopedia. New York, McGraw-Hill, 1969. 605 p.

Harrod, Leonard Montague. Librarians' Glossary...and Reference Book. 3d rev. ed. New York, Seminar Press, 1971. 784 p.

Library Association. Media Cataloguing Rules Committee. Non-book Materials Cataloguing Rules; Integrated Code of Practice and Draft Revision of the Anglo-American Cataloguing Rules British Text Part III. London, National Council for Educational Technology with the Library Association, 1973. 129 p.

Meek, Chester L. Glossary of Computing Terminology. New York, CCM Information Corp., 1972. 372 p.

Rodgers, Harold A. Funk & Wagnalls Dictionary of Data Processing Terms. New York, Funk & Wagnalls, 1970. 151 p.

Sippl, Charles J. Computer Dictionary and Handbook. Indianapolis, H. W. Sams, 1966. 766 p.

Sippl, Charles J. and Charles P. Sippl. Computer Dictionary and Handbook. 2d ed. Indianapolis, H. W. Sams, 1972. 778 p.

Verona, Eva. Statement of Principles Adopted at the International Conference on Cataloguing Principles, Paris, October, 1961, annotated edition with commentary and examples. London, IFLA Committee on Cataloguing, 1971. 119 p.

Weihs, Jean Riddle, Shirley Lewis, and Janet Macdonald. Nonbook Materials: The Organization of Integrated Collections. 1st ed. Ottawa, Canadian Library Association, 1973. 107 p.

Weik, Martin H. Standard Dictionary of Computers and Information Processing. New York, Hayden Book Co., 1969. 326 p.

Working Group on the International Standard Bibliographic Description. International Standard Bibliographic Description, Monographs. London, IFLA Committee on Cataloguing, 1971. 30 p.